ISSN 1749-8767

## DCC | Digital Curation Centre Case Studies and Interviews

# PrestoSpace: Preservation towards storage and access. Standardised Practices for Audiovisual Contents in Europe (http://www.prestospace.org)

---

Martin Donnelly, Victoria Boyd, and Jill Spellman
HATII, University of Glasgow,
Glasgow G12 8QQ
http://www.hatii.arts.gla.ac.uk

March 2008

Version 1.0

**Catalogue Entry**

| | |
|---|---|
| **Title** | DCC Case Study – PrestoSpace: Preservation towards storage and access. Standardised Practices for Audiovisual Contents in Europe |
| **Creator** | Martin Donnelly, Victoria Boyd, Jill Spellman (author) |
| **Subject** | Data curation; formats, processes and issues; system development; standards; legal factors; methodology, and problems overcome; human factors |
| **Description** | Explicit strategies are needed to manage 'mixed' audio visual (AV) archives that contain both analogue and digital materials. The PrestoSpace Project brings together industry leaders, research institutes, and other stakeholders at a European level, to provide products and services for effective automated preservation and access solutions for diverse AV collections. The Project's main objective is to develop and promote flexible, integrated and affordable services for AV preservation, restoration, and storage with a view to enabling migration to digital formats in AV archives. |
| **Publisher** | HATII, University of Glasgow; University of Edinburgh; UKOLN, University of Bath; Council for the Central Laboratory of the Research Councils. |
| **Contributor** | Joy Davidson and Andrew McHugh (editors) |
| **Date** | 18 March 2008 (creation) |
| **Type** | Text |
| **Format** | Adobe Portable Document Format v.1.3 (Acrobat 4.x) |
| **Resource Identifier** | ISSN 1749-8767 |
| **Language** | English |
| **Rights** | © HATII, University of Glasgow |

**Citation Guidelines**
Donnelly M, Victoria Boyd, Jill Spellman (March 2008), " PrestoSpace: Preservation towards storage and access. Standardised Practices for Audiovisual Contents in Europe ", *Digital Curation Centre Case Studies and Interviews*, Joy Davidson and Andrew McHugh (editors), Retrieved <date>, from http://www.dcc.ac.uk/resource/case-studies/prestospace

### *About the DCC*

The JISC-funded Digital Curation Centre (DCC) provides a focus on research into digital curation expertise and best practice for the storage, management and preservation of digital information to enable its use and re-use over time. The project represents a collaboration between the University of Edinburgh, the University of Glasgow through HATII, UKOLN at the University of Bath, and the Council of the Central Laboratory of the Research Councils (CCLRC). The DCC relies heavily on active participation and feedback from all stakeholder communities. For more information, please visit www.dcc.ac.uk. The DCC is not itself a data repository, nor does it attempt to impose policies and practices of one branch of scholarship upon another. Rather, based on insight from a vibrant research programme that addresses wider issues of data curation and long-term preservation, it will develop and offer programmes of outreach and practical services to assist those who face digital curation challenges. It also seeks to complement and contribute towards the efforts of related organisations, rather than duplicate services.

### DCC Case Study and Interviews

**Editors**

**Joy Davidson**
Humanities Advanced Technology and Information Institute, University of Glasgow (UK)
**Andrew McHugh**
Humanities Advanced Technology and Information Institute, University of Glasgow (UK)

# *Case Study: PrestoSpace: Preservation towards storage and access. Standardised Practices for Audiovisual Contents in Europe (http://www.prestospace.org)*

*This Digital Curation Centre case study is the result of an email questionnaire completed by Matthew Addis of IT Innovation in October 2005, an interview with Richard Wright of BBC Archives in September/October 2004, and subsequent email communications (2005-2008) between the PrestoSpace partners, Martin Donnelly, Victoria Boyd, and Jill Spellman of the DCC and HATII, University of Glasgow.*[1]

## Table of Contents

Executive Summary .....5
Introduction .....6
Objective .....6
Type of data curated .....7
Volume and costs of data curated.....8
Resolving the data volume problem.....9
Methodology, and Problems overcome.....11
Cost Modelling .....12
Human factors.....14
Evaluation .....14
Future Work and Considerations .....15
Annex I – PrestoSpace Objectives.....16
Annex II – Strategic Planning process .....17
Annex III – Technological issues.....19

[1] The answers given are a partial view of PrestoSpace by IT Innovation and only cover some aspects of The Project. Further information can be found in the deliverables available from The Project website, in particular the user needs deliverable D2.1 Other members of the PrestoSpace consortium may be willing to provide further input and can be reached through The Project coordinator. The project website is: www.prestospace.org

## Executive Summary

Explicit strategies are needed to manage 'mixed' audio visual (AV) archives that contain both analogue and digital materials. The PrestoSpace Project brings together industry leaders, research institutes, and other stakeholders at a European level, to provide products and services for effective automated preservation and access solutions for diverse AV collections. The Project's main objective is to develop and promote flexible, integrated and affordable services for AV preservation, restoration, and storage with a view to enabling migration to digital formats in AV archives.

'There is an issue, which is difficult to quantify: the value of data. However audiovisual media certainly contains some high value information, whose 'curation' is deserving of anyone's interest.'

Richard Wright, Technology Manager, BBC Archives[2]

**Introduction**

Contemporary audiovisual (AV) archives are often home to a mix of formats and media types, containing both analogue and digital content. Subsequently, AV archives face a number of unique challenges. Their holdings may be only partially catalogued, in poor condition, and/or held on obsolete media, thus hampering access. UNESCO estimates that over 200 million hours of broadcast footage is held in broadcast archives worldwide, "much of which is at risk of being lost and hence is the subject of international appeals for preservation," particularly in developing countries.[3] Explicit strategies are needed *now* to manage these 'mixed' archives.

The PrestoSpace Project brings together industry leaders, research institutes, and other stakeholders at a European level, to provide products and services for effective automated preservation and access solutions for diverse AV collections. The Project's main objective is to develop and promote flexible, integrated and affordable services for AV preservation, restoration, and storage with a view to enabling migration to digital formats in AV archives.

A typical broadcast archive may house many petabytes (PB) of digital AV content. According to Matthew Addis of PrestoSpace partner IT Innovation, this makes the Project's key driver simple: despite the newness of its creation, unless something is done soon—and on a large scale—substantial amounts of Europe's AV cultural heritage content will be lost forever. There is a clear need for new services that can digitise very large volumes of material at significantly lower costs than can currently be achieved.

**Objective**

About 25% of the BBC archive is accessed (taken off the shelf) per year. At the moment the BBC archive issues on average 155,000 items per month. The material can be used for a variety of purposes: preview, research, clip, selection, repeat, etc.[4]

Work on PrestoSpace began in February 2004, and was originally scheduled to run until June 2007. Project work was extended to January 2008 and will conclude with an all-partners public facing meeting in Rome. The Project comprises thirty-seven partner organisations across Europe, including major archives such as the British Broadcasting Corporation (BBC), Institut National de l'Audiovisuel (INA), Radio Audizioni Italiane (RAI), Österreichischer Rundfunk (ORF), and B&G (the audiovisual archive for The Netherlands). The Project has €9M support from the European Commission under the FP6 IST programme.

PrestoSpace follows on from the PRESTO project, which demonstrated that the cost of preservation work could be halved by applying a semi-automated 'assembly-line' approach.[5] PrestoSpace, in turn, aims to create

[2] From e-mail correspondence, 04/10/05

[3] FIAT/IFTA International Appeal for the Preservation of the World Audiovisual Heritage, http://www.fiatifta.org/aboutfiat/policy

[4] From interview with Richard Wright, Information and Archives, BBC, September/October 2004 http://www.dcc.ac.uk/resource/interviews/richard-wright/

[5] PRESTO's main finding was that "[t]he key to reducing costs without sacrificing quality is mass-transfers of material, using a 'preservation factory'. Software is needed to track all items moving through the factory, to eliminate manual effort wherever possible. The factory approach, compared to conventional item-by-item preservation work, reduces costs by AT LEAST 50% -- and some PRESTO partners have demonstrated reductions of up to 75%." http://presto.joanneum.ac.at/index.asp

blueprints for a standard preservation 'factory,'[6] providing affordable preservation services for the AV archive community, and assisting custodians in managing and distributing their assets. These factories will instantiate such services by preparing a generic business plan, contacting potential investors, and working with commercial partners to set up the services.[7]

The PrestoSpace philosophy is that access is the primary driver for preservation—that items have more value when they are accessible than when they are forgotten on shelves—and that it is the business priority based on this increased value, which drives the selection process.

The Project comprises four distinct work areas: (i) Preservation; (ii) Storage and Archive Management (SAM); (iii) Restoration; and (iv) Metadata Access and Delivery (MAD). Collectively, these areas investigate and develop the various facets of the 'factory' concept:

- **Preservation:** This area aims to develop automated video and audio preservation tools, which will be used in conjunction with manual quality assessment tools to assist in the quality-centred migration of AV content;

- **SAM:** This area develops models for the storage of digital archives, addressing the issue of safety for the long-term storage of digital archives[8];

- **Restoration:** This area is developing digital techniques that can be applied to digital audio and video content in order to remove defects. This will involve both hardware- and software-based techniques, combined into an overall AV restoration solution;

- **MAD:** This area deals with the cataloguing and indexing of digital content. This will involve a 'turnkey' system for the storage of, and access to, digitised content and associated metadata, including tools for semi-automated annotation and services allowing the search and retrieval of digital content via the Web.

We should note from the outset that PrestoSpace is not a digital management/curation system as such: it is not designed to support the full lifecycle of digital content created through preservation services. Instead, specific demonstrators are being constructed for component parts of the PrestoSpace 'factory,' and the additional services/tools that will operate in conjunction with it. (See Annex I, below, for a full list of The Project's deliverables.)

**Type of data curated**

Early in its development, PrestoSpace conducted a survey of European film and broadcast archives.[9] A great deal of information was gathered from twenty archives of varying sizes, including many state broadcasters. Main findings indicated a critical need for information on:

- new technology, related to storage devices;

[6] http://wiki.prestospace.org/pmwiki.php?n=Main.PresFact#Overview

[7] http://www.prestospace.org/project/objectives.en.html

[8] Currently, the benchmark standard for the construction of preservation environments is the Open Archival Information System (OAIS) Reference Model. This framework was a key conceptual framework for the Project throughout its progression. Project outcomes are complementary to an OAIS approach and should be mutually supportive. For more detailed information refer to *PrestoSpace Digital Repositories Explained*, http://www.prestospace.org/project/deliverables/D13-4.pdf

[9] The results of user surveys from PrestoSpace, TAPE (its sister project) and other sources can be found in PrestoSpace *Preservation Status: Annual Report(s) on Preservation Issues for European AudiovisualCollections,* http://www.prestospace.org/project/deliverables/D22-4_Report_on_Preservation_Issues_2004.pdf; http://www.prestospace.org/project/deliverables/D22-6.pdf; http://www.prestospace.org/project/deliverables/D22-8.pdf

- security and longevity of storage technology;
- costs and on standards.

Formats currently held by broadcast archives—as well as the target formats for ongoing and future digitisation projects—span portable digital file formats like MPEG-1, -2, -4, AVI, MP3 and WAV, in addition to formats that are linked to the medium on which the content is held, e.g. DigiBeta, D3 or DVCPro. The carrier medium is also a factor. Digital content may be stored on CD, DVD, tape, floppy discs, or a combination thereof, depending on the archiving strategy, the access need, and the content type, which continues to complicate the picture.[10]

**Volume and costs of data curated**

The user survey included questions on data volume, the findings of which are summarised below. These early findings represent predominantly analogue materials, but they demonstrate the scale of the digital archives that would need to be operated in the sector if all existing collections were to be digitised.

| Carrier | No. of cans/tapes |
|---|---|
| Film in Broadcast Archives (14 responses) | 4,238,857 |
| Film in Film Archives (5 responses) | 1,665,708 |
| Video tapes (16 responses) | 6,232,352 |
| Audio recordings (12 responses) | 5,321,301 |
| **Total AV items** | **17,458,218** |

These figures will also be augmented by new, 'born digital' material, the volumes of which can easily involve the same amount again in the case of active broadcast archives ingesting new content on a daily basis.

[10] For example, video might be held on tape masters at high resolution, on DVD for 'loan' copies, and transferred to discs at browse quality resolution for quicker Web access.

| Migration 2004-2006 | Items to Migrate | Transfer Service Capacity |
|---|---|---|
| Film | 29,168 | 17,500 |
| Video | 340,540 | 213,000 |
| Audio | 113,600 | 41,000 |
| **Total** | **483,308** | **271,500** |

The migration plans of the respondent archives reveal a need to transfer almost 500,000 items in two years. This may seem large, but when we note that 250,000 items per year represents only 1.5% of the 17,500,000 items held in these archives, the scale of the problem becomes clearer. Indeed, the current capacity for transfer is only slightly more than half of what is required. Furthermore, when it comes to costs, the available budget is less than half of that needed:[11]

| Film Video Audio Migration | Costs (€) | Available Budget (€) |
|---|---|---|
| **2004-2006** | **57,841,500** | **24,805,000** |

The most recent PrestoSpace annual report, *the Annual Report on Preservation Issues for European Audiovisual Collections (2007)*, examines "the cost of digitisation, and the long-term costs for 'digital preservation' *after* digitisation. [PrestoSpace] give(s) the costs, now, for digital storage—covering a range of storage options—and…forecast(s) the trend for costs over the next twenty years. The result is sufficient information for archives to create twenty-year plans for their new, digital archives. These plans are essential because digital materials, just like analogue ones, will not last forever and are just at much in need of maintenance as were the old 'tapes on shelves'. In fact, timescales for obsolescence of digital technology are, if anything, shorter than for the old materials. To balance the cost issues,

[11] See PrestoSpace Deliverable D2.1, especially pages 30-31, for a more detailed breakdown.

[PrestoSpace] summarise(s) current information on value: what it is and how to calculate it. *Public value* is especially relevant to heritage and broadcast collections, and there is growing awareness of the fact that the public value of audiovisual materials far exceeds previous narrow concepts of commercial value. Finally, the report provides an update on developments in online access, and new initiatives relevant to audiovisual materials—both political and from the audiovisual services (facilities) industry."[12]

**Resolving the data volume problem**

It should be clear there is neither sufficient capacity nor capital to transfer content at the rate planned by Europe's archives.[13] This problem is compounded by the fact that the transfer rates as currently planned are not high enough to avoid significant loss due to technical obsolescence, physical decay in addition to other factors such as inadequate intellectual and physical control over the AV archival holdings. At current rates, the best-case scenario is that 40% of tape-based content will be lost by 2045; the worst is that 70% will be lost by 2025.[14]

If we take the middle ground and assume that somewhere between 40% and 70% of tape content needs to go through PrestoSpace factories over the next thirty years or so, this is in the order of 2% of holdings each year. If we generalise this to film and audio as well, the factories will need to be capable of processing at least 300,000 items per year. But these estimates are generally accepted to fall short of the actual situation. Given that the initial PrestoSpace data survey only covered twenty archives, and there is likely to be at least as much content again that has not been counted, the real volume is likely to lie somewhere between 600,000 and 1,000,000 items per year. AV items average around twenty minutes in length, so the factories will need to process around 200,000 hours of material annually. Data volumes are also media-dependent: videotape, for example, is much more data-intensive than audio discs. If we assume for the moment that digital audio storage requirements are negligible compared to video, and that only DVD resolution is required for film and video (which is *not* the case), then the minimum data volumes are in the order of 800,000GB per year at DVD quality, and 20 times larger at video preservation quality = 16PB. Including significant amounts of film digitised at 2K or 4K resolution (instead of the 500 to 600 lines resolution of standard video) probably means another overall factor of two: 32PB per year [15]

Another way of looking at the data volumes is to observe that large AV archives with existing mass digitisation programmes have already accumulated multi-PB data volumes over the last five-to-ten years. The volume of data in need of digitisation across Europe is another order of magnitude higher, which brings us back to a figure of several PB per year. The challenge is clearly immense.

[12] To review the full annual report visit, http://www.prestospace.org/project/deliverables/D22-8.pdf

[13] The problem is, of course, not confined to Europe. In October 2005, the US Library of Congress announced its purchase of SAMMA, the System for the Automated Migration of Media Archives. SAMMA will migrate and digitise over half a million television items, as well as almost two million audio recordings. The Library calculated that it would take almost one hundred years to migrate and digitise these audiovisual collections manually.

[14] See PrestoSpace deliverable D22.4 for much more on this. URL: http://www.prestospace.org/project/deliverables/D22-4_Report_on_Preservation_Issues_2004.pdf

[15] This is a lot of digital data, but the estimates above also tend towards the conservative. Archival quality digitised video is nearer 80MB/s than the roughly 4 MB/s of commercial DVDs. If we look at the worst-case scenario for rate of loss (70% in 20 years), the data volume figure could easily be multiplied by a factor of 10. So PrestoSpace factories could conceivably be required to digitise at a rate of well over a PB each year, in addition to the volume processed by existing channels (e.g. in-house) or by direct interaction with service providers. Of course, PrestoSpace factories could facilitate this as well.

## System development

'In broadcasting, we've always complained that our formats were so short-lived, compared to film (nearly a century) and paper (many centuries). Video and audio formats were changing in less than 10 years. Now it looks like IT systems will continue to change significantly (to the point of needing replacement) in under five years. So the rate of 'churn' is going to be at least double, and media archivists generally are resigned to permanent migration.'

From interview with Richard Wright, Information and Archives, BBC, September/October 2004.[16]

Existing management systems (or *archive infrastructures*) range from *ad hoc* collections of media items on shelves with minimal analogue cataloguing, to large scale automated robot systems with fully integrated electronic catalogue and access mechanisms. Intermediate cases are more common, where very large archives maintain increasingly large digital collections in the same way as their analogue collections, e.g. as discrete items on shelves. Management strategies are equally varied, but tend to differ more in terms of their details rather than their overall goals. The exception is perhaps the fundamental difference in mission that often exists between broadcast archives and film archives.[17]

For broadcast archives, part of the process of providing 'access' typically means replacing analogue originals with (i) a high quality digital master copy, and (ii) additional, lower resolution 'browsing' copies. The latter will be accessed much more frequently, since low-resolution versions are sufficient for assessing the suitability of content. Access to the master *is* occasionally required; for example, when content needs to be repurposed for a new broadcast. In broadcast archives, preservation is essentially about preserving access to content over time, regardless of the specific media used to store it. The strategy, therefore, tends to be one of ongoing migration, typically from analogue-to-digital in the first instance, and then subsequently from digital-to-digital (D-D). This contrasts with film archives, which are more concerned with preserving the original object for as long as possible, even if this restricts access, e.g. by storing it in a freezer for 100 years. Film archives tend not to consider digitisation to be a viable approach to preservation, since the original carrier is often discarded, as are the devices needed to play the original media—and thereby recreate the original viewing and listening experience, which is problematised by digitisation. However, film archives do employ digitisation for access purposes, e.g. creating copies for education and public access via the Web, or for sale on DVD. This can, in turn, raise the need for a strategy to manage such digital surrogates.[18]

## Standards and Legal factors

PrestoSpace faces a major problem in the lack of standardisation across the sector. There is a real need to engage closely with industry and training bodies in order to improve the situation. The benefits of standardisation are well documented elsewhere, but a key encouragement

[16] *Op. cit.*

[17] Film archives are clearly very different to broadcast archives, in that their holdings consist primarily of feature films (which rarely need to be repurposed), and their largest user is the film-viewing public, through film screenings. For broadcast archives, digitisation is the solution to preservation and access problems, while film archives consider digitisation more the step to enhanced access. Due to the difference in user-base, film archives seldom have the same large-scale digitisation policies as broadcast archives. They often depend more on the professionalism of telecine operators, and spend more time verifying results. Archives are interested in automation, but approach its use with film more cautiously.

[18] It is also worth noting that the lifetime of a digital object is almost invariably longer than the medium on which it is stored, due to media decay and obsolescence of file formats, carriers, and players. Migration is thus inevitable in order to maintain access, and most archives have ongoing migration built into their archive models.

for the broadcast sector will be in the potential for additional or improved revenue streams. Many archives plan to promote the sale/licensing of content by opening their catalogues for external use. The value of holdings would certainly increase if a common metadata exchange standard were embraced.

The selection process for AV content is often accompanied by complex rights issues. Rights clearance remains a major unresolved problem. The right to use archival content is a significant factor when determining what to preserve. It is important that the retention policy works *in conjunction with* a digitisation selection strategy, and to bear in mind that these are not the same thing.[19]

The PrestoSpace user survey found that 57% of the broadcast archives [surveyed] claim between 50% and 100% ownership of their collections, although several of the larger archives claim 35% or less. Half of the film archives surveyed claimed 90% ownership, the other half less than 30%. In countries where archival enterprises were once nationalised (e.g. those in Eastern Europe), such rights issues are yet to be resolved.

Most AV archives use databases to manage their rights information, with around half of these operating independently of the object cataloguing system. The majority of archives polled believe that rights examination should be implicit within the preservation process: a streamlined rights management process, working in harmony with preservation workflows, would make rights management easier and faster, promoting increased content use and boosting income from licensing.

[19] Ways in which these strategies differ, and what the implications of the retention policy might be for long-term curation, are not within the terms of the PrestoSpace project. These concepts are further explored within the AVATAR-M project, http://www.it-innovation.soton.ac.uk/projects/avatar-m/avatar-m

**Methodology, and Problems overcome**

AV archives face a number of problems when planning preservation projects. They typically do not know the contents of their archives in detail (factors such as carriers/media, physical condition, volume, programme content), and few archives have funding sufficient to preserve all of their holdings. They therefore need to balance many factors:

- How much content can be preserved?
- How much it will cost?
- How long will this work take?
- What should be given priority?
- How much content can be expected to be lost?

Larger archives tackle these issues by adopting a planning process. The first level—called *strategic planning*—is concerned with constructing the 'big picture,' wherein archive managers decide on the best course of action for the archive as a whole. The second level—*operational planning*—involves the creation of a detailed plan for the execution of a preservation project.

The first stage of planning for migration to a new digital archive is to create, from physical/technical and content perspectives, a map of the current archive. The sheer volume of items often means that statistical sampling is the only viable approach, with a statistical 'map' used to assess the urgency and cost of preservation. Due to the fact that the condition and content of individual items will not be known until inspection prior to transfer, the preservation workflow necessarily involves an assessment and sorting stage to determine the action to be taken on an item-by-item basis.[20] A flow diagram of this process is given at Illustration 1, below.

[20] See PrestoSpace report, *Automation of Difficult Media*, http://www.prestospace.org/project/deliverables/D5-3_Difficult_Media.pdf

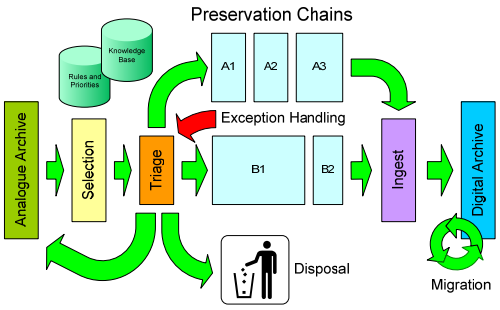


**Illustration 1: Workflow process diagram for preservation via migration and storage**

After auditing and quantifying the archival holdings, the procedures for ensuring the long-term preservation needs of the archive over time are derived, and the associated costs calculated. The difference between these costs and the budget available will influence the prioritisation of content, along with technical aspects (e.g. rate of decay, technical obsolescence) and the value and 'cost of loss,' or the potential financial losses incurred by the loss of material. This will lead to a year-by-year, high-level strategic preservation plan, subsequently extended into fuller details of operationalisation. (See Annex II for full details.)

## Cost Modelling

Efficient workflow operation is essential for minimising costs. This typically involves allocating items to one or more *transfer chains*, each of which is matched to a specific combination of carrier type and condition.[21] These transfer chains feed into the digital archive as part of the ingest phase. Discrete digital media suffer from many of the same problems as discrete analogue media, so to ensure no further loss the digital archive operates under an inbuilt and relatively short term (3–5 year) migration strategy, with interim refreshing of media (e.g. moving to higher capacity tapes) taking place more often. A specific timeframe for this is dependant upon the media and is intrinsic to the D-D migration cycle issue. Each individual technology standard and vendor has its own 'road map' for technical updates. [22]

The 'triage' of items prior to transfer is an essential stage in the migration process, and is underpinned by a preservation strategy developed through cost modelling. This is implemented as a set of rules to determine what action is to be taken, and is supported by a knowledge base that allows the cost of preserving each item to be estimated based on readily observable features, such as visible condition or chemical markers.[23] The two major factors that affect the cost of transferring items into a form that can be digitally archived are: (i) degradation (e.g. chemical decay, or 'wear and tear'); and (ii) technical obsolescence (the unavailability of media players, spare parts and/or skilled operators). There are no golden rules for calculating life expectancy due to decay; for example, tape condition will depend on factors including manufacturer, production batch, previous storage conditions, and frequency of use/handling.

The digital archive ingesting the items produced by migration can be modelled as a chain of mass digital media storage devices. The receipt and ingest of new content is continual, and this must be considered across successive storage solutions, as shown in Illustration 2, below. Items

[21] For example, tape may be in good condition, or it may be 'sticky' and in need of baking.

[22] The rapid and ongoing fall in cost of digital storage, together with its relatively short lifespan, means that it is sensible to purchase storage 'on demand' rather than as a single up front investment. PrestoSpace are modelling this approach to investigate this hypothesis.

[23] The knowledge base is built up from the known relevant parameters on the specified preservation chains. The need for and use of knowledge bases as part of planning and executing transfer projects is described within the PrestoSpace deliverables, *Automation of Difficult Media*, http://www.prestospace.org/project/deliverables/D5-3_Difficult_Media.pdf, and *Report on video and audio tape deterioration mechanisms*, http://www.prestospace.org/project/deliverables/D6-1.pdf.

ingested to the digital archive are assigned to the storage solution most appropriate to their type.[24]

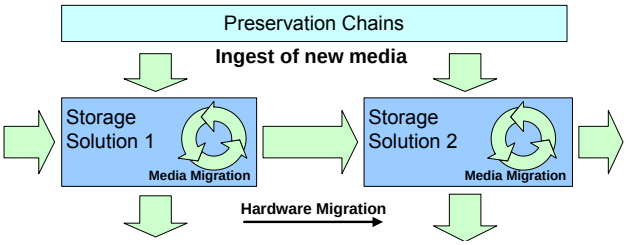


**Illustration 2: Time-based ingest, migration and disposal between storage solutions**

The relative cost effectiveness of different solutions will also vary over time, allowing, for example, transition from tape-based to disc-based solutions. The combined profile of ingest and migration for each media type determines the storage needed and cost as a function of time. Combining this with the transfer plan produces an overall projection of cost and loss on a yearly basis which can be optimised by investigating possible tradeoffs, e.g. the numbers/proportions of items that can be transferred, the costs involved in transfer and storage, time constraints, quality considerations, acceptable levels of loss, and the anticipation of future storage requirements.

Time-dependant factors (such as technology obsolescence and monetary inflation) will affect the cost of transfer or repair. The obsolescence lifecycle management of a particular tape player, for example, can be modelled as a cost multiplying factor which changes each year—costs are low when the player is still being produced, higher when it is initially discontinued, higher still when the manufacturer withdraws support, and highest when the second hand market is ultimately depleted of players.

[24] For example, CD quality audio is likely be assigned to disc, while high quality video at 80Mbit/s will more likely be assigned to tape.

The PrestoSpace model "includes a map of ingested items to digital media types and a matrix describing migration between different media and hardware solutions based on estimated life expectancy of storage solutions. The combined profile of ingest and migration for each media type determines the storage needed and cost as a function of time. Combining this with the transfer plan produces an overall projection of cost and loss on a yearly basis which can be optimised using a spreadsheet to investigate tradeoffs."[25] This provides a projection of long term transfer and storage needs, as shown in Illustration 3, below.

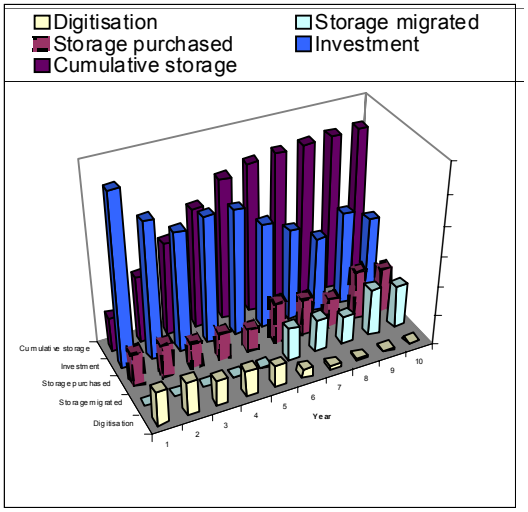


**Illustration 3 Transfer and storage projection**

PrestoSpace proposes a 'cost of quality' concept, which would provide archives with the tools necessary for planning migration using cost figures reflecting a balanced trade-off between quality and quantity. An example of this would be choosing between the cost of preserving an entire archive at 95% quality, or half an archive at 99% quality. In the PrestoSpace user survey, 68% of broadcast archives found this approach acceptable, and 81% said

[25] "*Planning the digitisation, storage and access of large scale audiovisual archives*," Addis, M., F. Choi, A. Miller, URL: http://eprints.ecs.soton.ac.uk/12231/01/045-poster.pdf

they would use price/quality arbitration tools from PrestoSpace in establishing their migration plans. Around two-thirds of film archives agreed with the cost of quality approach; 40% of these would use price/quality arbitration tools.[26]

**Human factors**

"Audiovisual archiving is a complex and multi-disciplinary domain spanning such diverse fields as chemistry, physics, signal processing, robotics, artificial intelligence and semantic interpretation. The challenge is to integrate partners of all domains representing the variety of competencies needed. The Project therefore brings together participants including archive owners, research centres from archive institutions, general research centres and universities, industrials, and international non-profit institutions.

The main organisers and managers of the Project are the seven core partners composing the Steering Board. The core partners are audiovisual archive owners, research and integration institutions and academic representatives. No industrial is present in the core group since the diversity of the technological solutions addressed by the Project does not allow singling out a common representative for the whole domain.[27]

External expertise for the Project was sought in two key areas. A Storage Technology Advisory Group (STAG), made up of various members of the digital storage industry, provided a solid knowledge base, while being made aware of the archive market's needs. A User Group (UG) of archives was created to ensure the key deliverables of PrestoSpace were delivered where needed. Both the STAG and the UG were to hold meetings and mini conferences throughout the life of the Project.

PrestoSpace also had a successful training component, which focused on the skills and techniques needed to access old and obsolete technologies rather than on the new skills required for the transition to digital archiving. Training materials produced comprise a series of DVDs, with subtitles in three languages, covering Audio, Film, and several different professional video formats.[28] To an extent, all the tools developed by the Project were for specific roles, and included notional user guides. In the SAM area, for example, information pertaining to the storage of assets in a digital form was provided in terms that would be familiar to a specific group of professional librarians.

**Evaluation**

In terms of formative evaluation, all the broadcast archives and service providers surveyed, and 60% of the film archives, consider the preservation factory approach to be compatible with their organisation/business orientation. Enthusiasm for this approach may be due in part to in its detailed process planning, and the tools that PrestoSpace is developing.

The business model and workflow for the final factory concept will depend to a significant extent on how the factory integrates with existing AV archives and their service providers. With respect to the migration of AV materials to the digital domain, it has been calculated that the factory approach is around 50% cheaper than the on-demand approach.[29] Larger archives with existing corporate archiving infrastructures are more likely to want to import digitised material and the corresponding

[26] Outsourcing is a further cost-reducing option for many archives, and an additional factor in PrestoSpace's cost modelling. 71% of the broadcast archives surveyed, and 100% of the film archives say they would be willing to outsource part of their planned collection migration.

[27] http://www.prestospace.org/project/consortium.en.html

[28] http://www.prestospace.org/training/index.en.html

[29] "The history of the 'Preservation Factory' concept," http://www.prestospace.org/PF_concept.pdf

documentation from the PrestoSpace platform to their own. Standardisation will be vital in order to facilitate this exchange, while smaller archives will be more likely to use the turnkey system for content management, incorporating new holdings and documentation, and making it accessible from outside.

Full integration and consistency of the Project will be guaranteed on three levels: first and most prominently by the core partners, who will participate in more than one work area or transversal work package. Second, associated partners will work in more than one work area or work package, if needed for the Project. For instance, the industrial partners will also contribute to system architecture, user requirements and to exploitation wherever appropriate. Details on that may be seen from the table on work distribution to work packages. Finally, the user requirements will be built on the experience and input of the UG.[30]

**Future Work and Considerations**

Perhaps the most important point that PrestoSpace aims to prove is that, in order to avoid the problems encountered to date with analogue archives of discrete media items on shelves, a well-managed, ongoing migration strategy is essential for the digital archive. In terms of time and cost savings, this is often best implemented through the use of automated systems, e.g., the use of mass storage devices rather than the continued reliance on discrete digital media.[31] This is based on a belief that cost modelling can assist with identifying the best strategy for migration, facilitating the calculation of year-on-year costs (and likely content loss), allowing decisions relating to trade-offs to be made in an informed way.

Complementary to the work of PrestoSpace is its sister project, Training for Audiovisual Preservation in Europe (TAPE).[32] It aims to contribute to action plans for preservation of the AV heritage by laying the groundwork for a programme for awareness raising and training of non-specialists. TAPE will bring together experts from large organizations and those involved in technologically advanced programmes (such as PrestoSpace and FIRST, funded under the IST programme of FP6) to develop a programme for training and supporting materials.

[30] http://www.prestospace.org/project/consortium.en.html

[31] The amount of AV material already committed to mass storage is relatively small, indicating that most archives have yet to make major investments. This raises the need for the targeted provision of information on storage technologies.

[32] http://www.tape-online.net/

***Annex I — PrestoSpace Objectives***

**Overall Objective: AV Preservation Services**

The main deliverable of the PrestoSpace Project is to develop and launch actual facilities and services for AV preservation. The Project will start these preservation factories by preparing the business plan, contacting potential investors and working with commercial partners to set up the actual services. These services will exploit the technological and industrial results of the Project. A strong economic factor supporting these preservation factories is a commitment by the major PrestoSpace consortium members to use these services.

**Deliverables**

*Preservation*

- A fast, affordable datacine[33]
- A contactless playback tool for audio disks
- An automated audio preservation tool
- An automated video preservation tool
- A manual tape condition assessment tool
- An information system for preservation management

*Restoration*

- A restoration management tool
- A defect analysis and description infrastructure
- A disk-to-disk, real-time restoration tool
- A digital film restoration software tool
- A set of high-level restoration algorithms

*Storage and Archive Management*

- A Web-guide and software tool for planning of storage for AV preservation
- A guide and software tool for business-case planning for AV preservation
- A logistics and quality insurance system for AV preservation

*Metadata, Delivery and Access*

- A semi-automatic description tool
- An export system for delivering preservation results to medium and large archives
- A turnkey system for delivering preservation results to small archives

[33] A specialised machine used in high-resolution film transfers.

## Annex II — Strategic Planning process

1. Audit and quantify archive holdings (volumes, content, carriers, condition). Questions to answer here include:

    a. What is the size of the archive, including the volumes for each carrier (e.g. 16mm film, 2″ video tape, ¼″ audio)? This defines scale of preservation needed.

    b. Are there unique or duplicate copies for the items in the archive? This is a factor in the risk of losing an item.

    c. What are the manufacturers, manufacturing batches, condition and age of the carriers (e.g. level 3 vinegar syndrome)? What are the storage conditions? This puts deadlines on preservation due to decay.

    d. What is the content on each carrier? This is needed for working out priorities for preservation (along with condition and technical obsolescence).

2. Estimate what needs to be done as a function of time to preserve the archive. This is based on carrier conditions and other problems such as anticipated technical obsolescence of equipment needed for transfer, e.g. 2″ or ¾″ tape players. The output of this task is the long-term preservation need of the archive over time.

3. Estimate the total costs of preserving all holdings (film, audio, video). Global costs are essential in quantifying the overall problem. There is limited need for accuracy at this stage, since costs will vary with time (year to year). Costs should also be broken down into carrier types, as well as being based on in-house facilities and market conditions for outsourcing, e.g. approx 60Euro per hour for ¾″ tape.

4. Estimate the costs for archive preservation as a function of time. Ensure that this respects deadlines (e.g. due to decay of carriers). Use projection techniques (if possible) to take into account trends identified from past experience, e.g. falling cost of using third-party preservation services. Estimate costs over time for each carrier since these have different time constraints (e.g. vinegar syndrome for acetate film) and can have dramatically different costs for transfer.

5. Compare costs against budget (over time) to determine whether budget is sufficient, or, if not, to identify what budget is needed.

6. Prioritise preservation based on both content (value and ‘cost of loss’) and technical aspects (rate of decay, technical obsolescence, etc.). Note: this is not the same as selection since no content is excluded from the preservation plan at this stage; instead everything is included and is given a relative importance. (Selection actually happens naturally since decay or obsolescence will eventually destroy anything that is low priority since it won’t be preserved soon enough).

7. Estimate what can be achieved over time (year by year) based on priorities and available budget. This is the high-level preservation plan and shows what will happen to the archive over time.

The strategic level process is then extended into operation and implementation of the archive preservation plan:

1. Create detailed yearly plan (costs and volumes) broken down by carrier (e.g. 2″, 1″, ¾″ tape) by task (e.g. assessment, cleaning, transfer, quality control etc.) and by service provider (in house or subcontractors). Note that in the detailed plan, tasks can be very specific and have differing levels of difficulty (e.g. bar coding, attaching tape leaders, tape re-splicing).

2. Create 'queues' of preservation jobs based on priorities. Typically, a queue will be set up for each carrier (e.g. 2″ tape, 35 mm film, ¼″ audio, etc.)

3. Create and operate a preservation chain for each queue. A chain consists of cleaning, transfer, quality control, creation of digital surrogates for different applications (e.g. browsing copy), storage and indexing, access etc. The chain might also include restoration. Parts of the chain might be outsourced. Outsourcing may involve using multiple providers to achieve required throughput, spread the risk, and allow load balancing.

4. Continuously monitor preservation progress against targets and budget. Refine predictions on a regular basis to update detailed yearly plan.

5. Periodically re-evaluate overall plan, in particular on the basis of (re)assessment of progress, archive holdings and condition, technical obsolescence, and changes in cost for preservation activities. This reveals what has been done and for what cost as well as what still needs to be done and what it is expected to cost.

### *Annex III — Technological issues*

The number of technologies employed by PrestoSpace is enormous, and too long to list here.

Full details—including reports on documentation standards, multimedia content-based analysis, technical manuals for analogue tools, and various comments on technologies—can be found in the publicly accessible deliverables, and in the annual report on the state of Europe's archives, available from the PrestoSpace website.

The SAM website provides a further resource for those interested in technologies for digital storage.[34]

[34] http://prestospace-sam.ssl.co.uk/index.html